

The aggregation problem in its hystorical perspective: a summary overview

WYE CITY GROUP

Global Conference on Agricultural and Rural Household Statistics

By **GIANCARLO LUTERO**

Presented by **EDOARDO PIZZOLI**

ISTAT, Roma, Italy

Outlines

- Introduction
- Longitudinal aggregation
- Aggregation bias
- Temporal aggregation
- Contemporaneous aggregation
- Composite indexes
- Concluding remarks

Introduction

- 1. Longitudinal or spatial aggregation**, aggregation through geographic space (usually Regions or lower administrative areas) or a combination of physical and institutional units (households, businesses and institutional sectors), or the productive sectors;
- 2. Temporal aggregation**, passage of time series from an higher frequency data (i.e. quarterly) to lower one;
- 3. Contemporaneous aggregation**, aggregation is done across variables (i.e. the construction of a composite index such as the Human Development Index), or between different prediction models (i.e. the combining or pooling forecasting)

Introduction

Available statistical and econometric procedures (parametric and non-parametric) to address the problem of aggregation may be classified into the following macro-groups:

- stochastic framework based on regression models:
 - static models (cross section data);
 - dynamic models (time indexed, typically ARIMA models or OECD Composite leading indicators);
- deterministic framework, non-model based (composite indexes, social indexes).

Advantages to work with aggregated data

- **Data availability:** choice between aggregated data and a disaggregated ones is imposed by availability of figures;
- **Simplicity and parsimony:** few numbers is very effective in summarising complex problems in a simple and comprehensible way;
- **Faster statistical elaboration:** this is fundamental when data should be released at fixed deadlines;
- **Incomplete or unreliable statistical sources:** macro-data, in some cases, can solve the problems in micro-data, as well as reduce their impact on model specification, that is the likely introduction of a systematic error;
- **New surveys are very expensive:** it is needed to evaluate their real contribution in terms of statistical information.

Advantages to work with disaggregated data

- **Informative completeness:** it is preserved all informative content and data reduction (aggregation) is only one of possible several choices for analysis purposes;
- **Better estimates:** greater available number of observations involves more degrees of freedom in data treatment, and, subsequently, more efficient estimates, more powerful statistical tests, etc.
- **Model specification heterogeneity:** microdata availability agrees to consider various model specifications across micro units, to suit the circumstances;
- **Better predictions:** it is a consequence of a better and differentiated micro specifications.

Aggregation bias

Aggregation bias concerns possibility that, relating to true β value, asymptotic bias may be greater in the aggregated estimator rather than in disaggregated one, that is

$$\left| \text{plim } \hat{\beta}_d - \beta \right| < \left| \text{plim } \hat{\beta}_a - \beta \right|$$

Aggregation bias

Perfect aggregation

- **Micro homogeneity:** all microparameters are equal to aggregated one, so they are constant over the longitudinal dimension;
- **Compositional stability:** it regards joint probability distribution of regressors and it occurs when composition of the regressors across units do not change over time.
- **Symmetric distribution:** regressors are white noises.

Longitudinal aggregation

Aggregation bias in static or dynamic aggregated stochastic models could be determined by

- Moving from micro parameters values to macro one;
- misspecification of equations (omitted variables in micro models or, conversely, models overparametrization);
- measurements errors (microdata);
- missclassification errors in the process of aggregation;
- dynamic interrelation between the behavioral structural micro equations.

Temporal aggregation

- the order of AR(p) component is invariant to temporal aggregation, *unless* that in some particular circumstances;
- Otherwise, the AR coefficients are very different relating to the originals;
- Furthermore, MA component is generally introduced as consequences of aggregation while was absent in original processes;

Temporal aggregation

- an ARMA process is generally introduced by aggregation of not-ARMA or by several ARMA processes;
- in most cases the aggregated stochastic process is more complicated respect to the primary;
- it is useful to stress that, however, disaggregated models are information richer and number of observations is k -times greater than aggregated ones, although parameters contain all information about disaggregated sample, with all the consequences in terms of consistency and efficiency of estimates.

Contemporaneous aggregation

It regards combination of several indicators, relating to a target variable.

- “Economistic” and unidimensional approach cannot be considered satisfactory to analyze a complex phenomena like society well-being or rural development
- Adoption of social indicators is justified by their power to provide additional useful information with respect to GDP per capita or other main official statistics

Contemporaneous aggregation

Saisana and Tarantola (2005) report positive features on employ of a composite index:

- it summarizes complex and multidimensional issues;
- it is easier to interpret respect to many different indicators;
- it facilitates communication with public opinion (citizens and media);
- it may assess progress of specific country over time.

Composite indexes

Construction of a composite indicator may be synthesized in the following steps:

- Strong theoretical references;
- Variables selection;
- Missing data imputation;
- Multivariate analysis approach;
- Data normalisation;
- Weighting scheme identification;
- Aggregation of indicator variables;
- Check over composite indicators robustness and sensitivity analysis.

System of weights

- **Subjective:** weights are attributed following a subjective probabilities (degree of confidence) that certain events occur;
- **Objective:** system of weights result from application of a statistical rules or methods
 - **exogeneous**, pre-determined: weights are imposed by nature of problem and consequently they are defined in advance;
 - **endogeneous:** system of weights is estimated by means a constrained optimization procedure;
- **equal weighting:** it has been also considered composite indexes that assign an equal system of weights; this is a solution to reduce interferences to a minimum, or it is a consequence of lack of information.

System of weights

- **static weights:** system of weights do not change over time because analysis is involved with evolution of variables;
- **dynamic weights:** this is best solution when weights are output of a parametric optimization procedure.

Aggregation methods

- *Linear additive aggregation;*
- Geometric aggregation;
- Non-compensatory multi-criteria aggregation;

Concluding remarks

- Choice between macrodata and microdata and, after that, between aggregated or disaggregated data, cannot be resolved by *a priori* (by theory) reasoning, but it may find only an empirical solution, in context of particular applications, because several elements (estimation uncertainty, models misspecification in some level of aggregation, endogeneous dynamic relation between predictors) could be sources of statistical bias:
- As suggest Blundell and Stoker (2005) there are two good reasons to face aggregation topic with a new attitude: 1) the increasing availability of data on elementary units over sequential time periods and 2) the more and more rapid rise in computing power;

Concluding remarks

- Increase of data accessibility allowed discovery and application of relative new methodology like Dynamic Factor models or Panel Data models. These are two of potential partial solution for bias resulting from longitudinal and temporal aggregation because are considered, in the same model, statistical measures typical of individual events and aggregated behaviours.
- a satisfactory aggregation process, usually necessary for analysis on extended areas or large groups, should be able to suggest the use of statistical techniques, a weighting system or a number of statistics that reshapes the economic and social representation in a unbiased way and should not induce final users (policy-makers, public opinion, etc.) toward misleading evaluations and decisions.

Barker T. and Pesaran M.H. (1990), "*Disaggregation in econometric modelling*", Routledge, London and New York.

Forni M. and Lippi M. (1997), "*Aggregation and the microfoundations of dynamic macroeconomics*", Oxford Clarendon Press.

Forni M., Hallin M., Lippi M., Reichlin L. (2000), "*The generalized dynamic factor model: identification and estimation*", *The Review of Economics and Statistics* vol. 82, pp 540-

Grunfeld Y. and Griliches Z. (1960), "*Is aggregation necessarily bad?*", *The Review of Economic and Statistics*, vol. 17, no. 1.

Hoffman, A., Giovannini E., Nardo M., Saisana M., Saltelli A. and Tarantola S. (2008), "*Handbook on Constructing Composite Indicators: Methodology and User Guide*", OECD Statistics Working Paper, Paris.

Lee K.C., Pesaran M.C. and Pierson R.G. (1990), "*Testing for aggregation bias in linear models*", *The Economic Journal* (Supplement), vol. 100, pp 137-150.

References

- Orcutt G. H., Watts H. W. and Edwards J. B (1968), "*Data aggregation and information loss*", American Economic Review, vol. 58, no. 4.
- Pesaran M.C., Pierse R.G. and Kumar M.S. (1989), "*Econometric analysis in the context of linear prediction models*", Econometrica, vol.57, no.4, pp 861-888.
- Saisana M. and Tarantola S. (2002), "*State-of-the-art report on current methodologies and practices for composite indicator development*", EUR 20408 EN, European Commission-JRC: Italy.
- Sen A.K. (1985), "*Commodities and capabilities*", Amsterdam, North Holland
- Stiglitz J.E., Sen A. and Fitoussi J.P. (2009), "*Report by the commission on the measurement of economic performance and social progress*", Paris, CMESP 2009.
- Theil H. (1954), "*Linear aggregation of economic relations*", Amsterdam, North Holland

References

Timmermann, A. (2006). “*Forecast combinations*”, in *Handbook of economic forecasting* (edited by Elliot G., Granger C. and Timmermann A.), vol. 1, pp. 135-196. Amsterdam, Elsevier.

United Nations (2001), “*Human development report*”, United Kingdom, Oxford University Press.

Zellner A. (1962), “*An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias*”, *Journal of the American Statistical Association*, vol. 57, no. 298.